

A MAC PROTOCOL AND A SCHEDULING APPROACH AS ELEMENTS OF A LOWER LAYERS ARCHITECTURE IN WIRELESS INDUSTRIAL LANS

Andreas Willig

Technical University Berlin, Telecommunication Network Group

Sekr. FT 5-2, Einsteinufer 25, 10587 Berlin

email: awillig@ee.tu-berlin.de, phone: 49 30 31423831, fax: 49 30 31422514

June 9, 2003

Abstract

This work-in-progress paper describes some elements of our efforts in defining an architecture and an appropriate set of protocols for industrial local area networks (ILAN) with wireless communication media. The major focus is on lower layer protocols. The main elements are a description of data link layer services, a MAC protocol and a scheduling approach taking into consideration special properties of wireless media. The proposed MAC protocol is a variant of TDMA attacking its known problems. The goal of the link state dependent scheduling approach is to increase transmission efficiency over wireless links.

1 INTRODUCTION

Industrial local area networks (ILANs) differ significantly from classical LANs like ethernet due to some special requirements of their application area, e.g. the need for hard timing and bandwidth guarantees and supporting priorities.

On the other hand wireless LANs have received a lot of attention in research and industry due to some attractive possible features like mobility and reduced cabling needs. These features can be very attractive for use in industrial environments.

My research is concerned with the definition and implementation of wireless industrial LANs. Such a system is built from a lot of different elements. In this work-of-progress paper I will present some important architecture and protocol elements of the lower protocol layers of wireless industrial LANs.

If one wants to use wireless media in industrial LANs, some special properties of wireless data transmission has to be kept in mind. These properties have a strong impact in the design of protocols for media access control (MAC), which in turn play a key role in establishing hard timing and bandwidth guarantees. In this paper, after describing the services needed in industrial LANs, I propose a TDMA based, but much more flexible MAC protocol, which is capable of supporting a wide range of packet scheduling policies, thus giving guarantees for worst case medium access time and worst case minimum bandwidth. It is possible to support priorities. I also pro-

pose an approach for the design of packet scheduling algorithms which take the nonstationary error characteristics of wireless media into account and aims to increase channel throughput. This approach is called *link state dependent scheduling* and is based on the observation, that the throughput increases significantly, if transmission on a link (the shared channel between each two stations is seen as a separate link) in bad state (i.e. with currently much errors) is postponed for a while, hoping that meanwhile the link turns into a good state (i.e. with currently few errors) [3].

This is a work-in-progress-paper. So I can only present some main ideas now. Numerical and analytical results or exact and verified protocol specifications are not yet available.

This paper is organized as follows: in the rest of this section some short remarks on related research are given. In the next section 2 a short description of the important properties of wireless transmission is given, followed by the basic assumptions about the whole systems structure of an industrial LAN using wireless technology and by a description of the lower layer protocol architecture. In section 3 the data link services needed for industrial applications are discussed. Section 4 describes FTDMA, the proposal for a MAC protocol, and section 5 describes the link state dependend scheduling approach, using the earliest-deadline-first- (EDF-) scheduling algorithm as an example. In section 6 the conclusions are presented.

1.1 Related Research

One group at the Ecole Polytechnique Federale de Lausanne (EPFL) is researching in the integration of wireless stations into a fieldbus using the FIP protocol [16] or using the DECT protocol [17]. With the FIP protocol the approach is to build a gateway which, on the wireless side, serves as a central base station using a TDMA protocol. The gateway mirrors the process variables produced by the mobiles and consumed by the fixed stations, and it also mirrors the variables produced by the fixed stations and consumed by the mobiles (for a short description of the FIP protocol see [6]).

In the OLCCHA project also the FIP protocol is used [13].

There are a lot of efforts in defining an architecture and suitable protocols for wireless ATM (e.g. [21], [14], [20], [12]), which can be attractive for use in wireless industrial LANs, although they have a slightly different service model. Some proposals use stochastic medium access protocols ([21], [12]), or a demand assignment scheme ([14], [20]) with contention occurring in reservations.

2 ASSUMPTIONS AND SYSTEM ARCHITECTURE

In this section some properties of currently available wireless technology are discussed, as they are relevant for the design of lower layer protocols and services. The overall system structure of an industrial LAN using wireless technology is presented, followed by a description of the lower layers protocol architecture for individual mobiles. This should help to locate the below described elements within the whole system.

2.1 Wireless LANs

The field of wireless LANs and wireless data communications is very large. Some general references are [19], [2], [8], [5].

Radio waves are used for information transmission. They can have different properties: if high frequencies are used (20-60 GHz), penetration of walls is not possible, in contrast to lower frequencies, e.g. the unlicensed 2.4 GHz ISM band (Industrial, Scientific and Medical band). In this paper I focus on the ISM Band.

For radio waves in the ISM band there are different sources of transmission errors:

- Fast Fading: radio waves in the ISM band can penetrate walls and human beings and can be reflected on different materials. As a consequence multipath fading occurs. The signal strength at the receiver can vary in a large scale due to interference of different signal paths. At a frequency of 2.4 GHz with

only two signal paths the distance between complete deletion and amplification of the signal due to interference is only 6.5 cm (equivalent to $\lambda/2$).

- Slow Fading: if a large obstacle, which can not be penetrated by the waves is situated between sender and receiver, the signal strength is reduced significantly. A signal must be reflected a number of times, before it can be received. Each reflection means also a loss of signal strength.
- Delay Spread: due to multipath fading signals of two consecutive symbols can overlap at the receiver. The result is called intersymbol interference.
- Noise: radio noise can be induced by other equipment, e.g. microwave ovens, heavy machinery and drives, portable telephones (Handys), data transmission in adjacent cells and so on.

Errors due to path loss are not considered, since in our case transmission is performed only on short ranges. The impact of the different error sources can vary over different time scales. For a moving station signal loss due to fast fading is expected to occur with a high frequency (up to 100 times per second if the mobile is moving fast [5]), while the frequency of change in signal strength due to slow fading is much lower (in the range of several seconds). For a more detailed and complete discussion of error sources and quantitative results see [11], [10], [4], [5], [19].

It is assumed, that the wireless technology has the following essential properties:

- the channel has a bandwidth of at least 4 MBit/sec and all stations share the channel (broadcast medium). The 2.4 GHz ISM band is used.
- it is not possible to transmit and receive simultaneously on the same channel (due to overcharge of the receiver filters). So it is very hard to implement collision detection in a MAC protocol.
- the time to perform the switching from sending to receiving mode is negligible. Only a short preamble acquisition time is needed (this assumption is optimistic with respect to currently available technology).
- Spread-spectrum transmission is used for reducing the error rate of the channel. I expect frequency hopping spread spectrum transmission to be more robust in the presence of multipath fading.

2.2 System Architecture

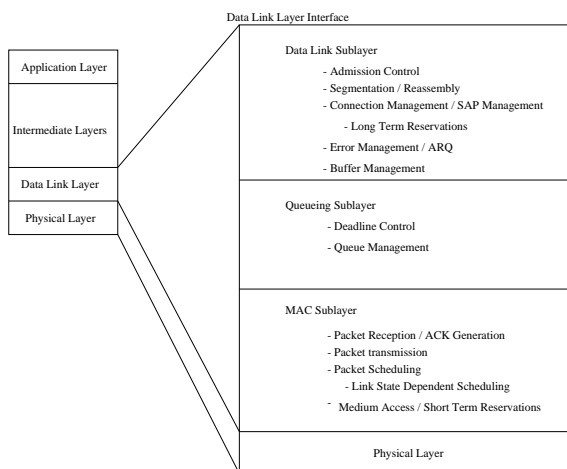
The industrial local area network (ILAN) we have in mind has the following overall structure: a wireless picocell consists of a central base station (BS) and one or more mobiles. Within a cell two mobiles communicate directly, i.e. on a peer-to-peer basis. Two mobiles in different cells use their respective BS and the backbone network interconnecting the base stations for forwarding packets. The size of a picocell is chosen in such way that all stations (mobiles and BS) within a cell can hear each other. Adjacent cells may use the same or different frequency bands; this choice has a strong impact on the necessary handover algorithm. The interference between adjacent cells has to be controlled by proper setting of transmit power at the sender or sensitiveness of the receiver (i.e. the required signal to noise ratio).

ATM is used as the backbone network [7]. Every base station is directly connected to an ATM switch. Communication between base stations is handled via fixed ATM connections. Instead of a dedicated ATM network also a larger ATM network can be used which is shared with other applications (e.g. office applications), but in this case fixed virtual paths with dedicated VPIs and a dedicated bandwidth and guaranteed delay should be used.

ATM is attractive due to its ability to support a large variety of Quality-of-Service (QoS) requirements in a very fine granularity. For an ATM connection a minimal bandwidth and a maximal end-to-end-delay can be guaranteed. It is possible to find a proper mapping of the industrial and hard-realtime QoS-requirements to the ATM-QoS-model.

2.3 Lower Layer Protocol Architecture

Within a mobile the lower layers are including the data link-sublayer, the queueing sublayer, the MAC sublayer and the physical layer (see following figure).



The data link sublayer performs connection management, error management (ARQ), segmentation and reassembly, and as very important part admission control. For every datagram or real-time connection it has to be

checked, whether related timing or bandwidth requirements can be guaranteed without affecting the guarantees for already existing connections or datagrams. However, under some circumstances (arrival of higher priority packets, see below) the deadline associated with a packet cannot be met, even if admission control was successful. So the packet must be removed from the queue in which it is stored. This is, beside queue management, performed in the queueing sublayer, which periodically checks for late packets. The MAC sublayer actually performs transmission and reception of packets, including generation of immediate acknowledgements. For packet transmission two important tasks need to be handled: the media access control algorithm determines, when packets may be sent, and the packet scheduler determines by applying a selection rule, which packets are to be sent next. For proper admission control in the data link layer the selection rule must be taken into account (e.g. for a simple earliest deadline first selection rule there exist a very simple admission control test for cyclic connections; see below).

One of the main ideas of this paper is an improvement of the packet scheduler: it may decide to postpone transmission of a packet for a short time if the link to the destination is currently bad. This short time must be selected in a way that deadlines are preserved. Instead another packet which uses a good link is transmitted. However, this scheme should be applied *transparently* for all upper layers.

3 DATA LINK SERVICES FOR INDUSTRIAL LANS

In this section I give an description of the data link layer services needed in an ILAN with its stringent timing- and QoS-requirements. This description is derived from a rough description of the traffic characteristics in industrial networks.

3.1 Characteristics of industrial Traffic

The term industrial traffic denotes the traffic generated by a wide class of applications. Some examples are production control, control of chemical plants, air control, communication systems in cars, planes and trains, power station control and so on. These applications are very complex, so their functionality needs to be distributed to a number of systems, which communicate with each other. The communication must be performed under hard real-time constraints, since missing of a deadline can lead to a catastrophe.

All these applications have in common, that for some data flows they need a bounded worst-case end-to-end delay, a guaranteed minimal bandwidth, sometimes a guaranteed maximal jitter, priorities and they can often not tolerate packet losses. These properties make industrial traffic different from best-effort-traffic (no losses and no

timing requirements) and multimedia traffic (hard timing requirements but relatively small loss sensitivity).

In a distributed control system many tasks are of a cyclic nature, e.g. drive control in a CNC. The quality of the control is determined by frequency, delay and the jitter of control messages. In many cases, e.g. for reading actual positions, it is essential, that the application knows the *freshness status* of the actual positions, i.e. how much time has passed since the last actualisation of the position, since for a fast moving axis this value is only valid for a short time.

Another important set of tasks is aperiodic or *sporadic*. Every control system needs to determine a set of exceptional and erroneous conditions in the system to be controlled. In these cases the controller must react immediately with performing appropriate actions and communicate the *alarm* to other controllers. It is often required, that the whole system falls into a secure emergency stop state, if some prespecified alarms or situations occur.

Following this, it is possible to classify industrial traffic due to its importance¹:

1. high priority sporadic messages, e.g. alarms, error messages, emergency stop messages
2. periodic messages
3. low priority sporadic messages (best effort)

Within a class further differences in importance are possible, e.g. messages for visualization purposes should be transmitted frequently with 25 Hz but they are not so important as messages needed for temperature control in a power plant control.

But it is very important to note, that the priority between these three classes must be reflected in the sequence of packet transmission: **a high priority sporadic message must always be transmitted before messages of other priority classes**. If the whole bandwidth is needed for high priority messages, no cyclic or best effort messages should be transmitted anymore. By the same principle cyclic messages have always priority over low priority messages. As a consequence, the maximum packet length must be tightly bounded, since an arriving high priority packet must be transmitted with as small delay as possible.

3.2 Service Specification

I propose three different services of the data link layer: an acknowledged datagram service, an unacknowledged datagram service and a connection oriented service, in which every packet may be acknowledged or not.

Acknowledged datagram service

The acknowledged datagram service is a connectionless service. Every packet must be ack'ed by the receiver

within a bounded time. Multicast or broadcast is not possible (who should send an ack?).

The user requests an acknowledged datagram with the `ACKDATA.request` service primitive. The reception of an acknowledged datagram at the receiver is indicated via the `ACKDATA.indication` service primitive. The user at the sender is informed about the result (proper acknowledgment or packet loss) using the `ACKDATA.confirmation` service primitive. If the sender does not receive an acknowledgement within a bounded time, the packet is retransmitted a number of times, which is upper bounded. With each packet a deadline can be associated.

The acknowledged datagram service should give the following guarantees to its users:

- if the packet deadline is not null then: after an `ACKDATA.request` the corresponding `ACKDATA.indication` occurs within the deadline or never.
- the time between `ACKDATA.request` and `ACKDATA.confirm` is bounded above by the deadline and a fixed short time for internal packet processing.
- the acknowledgement is transmitted immediately after successful reception of a packet.
- the maximum number of retransmissions is bounded by `Max_Retry`.
- if the packet is a high priority packet it will always be transmitted before low priority or cyclic packets.

Unacknowledged datagram service

The unacknowledged datagram service is also a connectionless service. The receiver sends no acknowledgement. Broadcast and Multicast can be used.

The user requests the unacknowledged datagram service with the `DATA.request` service primitive. The reception of an unacknowledged datagram at the receiver is indicated via the `DATA.indication` service primitive.

This service is very similar to the acknowledged datagram service, except for the following differences: 1) no `max_retry` parameter is given. There is always only one attempt to transmit a packet. 2) The receiver transmits no ack.

¹In the PROFIBUS system [9] the same classification is used, the classification in [1] is also very similar.

Connection-oriented service

The connection oriented service is the most complex. It is divided into a connection setup phase, a data transmission phase and a connection release phase. No concrete signalling protocols (e.g. two-way-handshake or three-way-handshake or whatever) are prespecified, but all protocols should be able to perform negotiation of QoS-parameters. The transmitted data can be acknowledged or not. The sender sets up a logical simplex channel to the receiver. If a minimum guaranteed QoS is required, the sender, the network and the receiver needs to reserve the needed resources, e.g. buffers, transmission slots, delay, bandwidth. The network (base station and backbone) and the receiver must decide, if the requested resources are available or not (call admission). If so, the new connection is established, otherwise it is rejected.

Some of the QoS-parameters needed for admission control at connection setup time are the following: maximum number of retransmissions (in case of acknowledged data), connection priority (relative to other cyclic connections), period (rate at which the user gives new packets for transmission), deadline for every packet, minimal size of data packets (which is guaranteed) and maximal size of data packets (which can be guaranteed or transmitted on a best effort basis, depending on the admission control scheme).

The bandwidth guarantees can be given under the constraint, that no high priority messages need the bandwidth, since these always have priority over cyclic messages.

The described parameters are used by the network and the receivers to determine, if the new connection is schedulable, i.e. if the requested guarantees can be given without affecting the guarantees of already existing connections (call admission). It depends on the implementation, if the call admission algorithm is performed only in the base station, in the intermediate switches and the receiver, or additionally in the mobile requesting the call.

With the given parameters a CBR (Constant Bit Rate) service can be described and implemented, that can guarantee bounds on delay and bandwidth.

For VBR services the same guarantees as in ATM can be expressed: a minimum bandwidth (and delay) is guaranteed, anything more is transmitted on a best effort basis, but always before low priority sporadic messages (statistical multiplexing). Since for call admission at least the minimal size and period parameters are needed, it may be a good idea to take also the maximum size parameter into consideration. The idea is to restrict the number of VBR sources with the hope of reaching a low probability that the bandwidth is temporarily exceeded by all the connections².

It is important to note that this services differ significantly from the ATM service model, however, a mapping of the ILAN services to ATM services is possible.

The main difference is, that in an industrial network high priority traffic can suppress all other types of traffic, i.e. cyclic traffic and low priority traffic, in a way that the whole available bandwidth is used by high priority traffic. This is not possible in ATM, since cyclic traffic (CBR) receives its bandwidth share independent of other sources. However, since ATM networks have a far greater bandwidth than currently available with wireless media, this property can be imitated by reserving CBR connections only for high priority messages with the bandwidth available in a wireless cell.

4 DESCRIPTION OF THE FTDMA-MAC-PROTOCOL

In this section the FTDMA (Flexible TDMA) MAC scheme is described. The TDMA scheme is attractive for industrial environments with hard timing constraints since all stations a minimum bandwidth and a maximum medium access delay can be *guaranteed*, regardless of the behaviour of other stations. The main disadvantage of TDMA is its lack of flexibility in the assignment of time slots, the fact that an unused slot cannot be used by other stations and a relatively large access delay even under low load. I propose FTDMA (Flexible TDMA), a modified TDMA scheme which overcomes these difficulties.

4.1 Frame structure

All mobiles and the base station share a common channel. Time is divided into frames of fixed length, denoted as t_f . A Frame is subdivided into a fixed number of slots for management and signalling purposes and a fixed number N of slots for data transmission. All the slots for data transmission have the same fixed length. In every data slot exactly one packet and the corresponding (short) immediate acknowledgement can be transmitted. The mobiles and the base station has enough buffer capacity to store at least N packets.

We assume two different types of mobiles: realtime-clients (rt-clients) are stations that need time and/or bandwidth guarantees for some of their communication tasks, while non-realtime-clients (nonrt-clients) do not need any guarantees at all. The number of rt-clients M per base station is bounded. The rt-clients register themselves at the base station and will be polled in every frame. In response to the polling they can request or release reservations. If they request a reservation the base station decides within one frame length whether the reservation can be accepted or not. The base station assigns the available slots first to rt-clients which have made reservations. The remaining slots (if available) are marked as free and all stations (rt-clients and nonrt-clients) may compete for access to the free slots using a contention based protocol (e.g. slotted ALOHA with a suited backoff algorithm).

²In [18] an overview of several VBR source characterizations and proposed admission control algorithms is given.

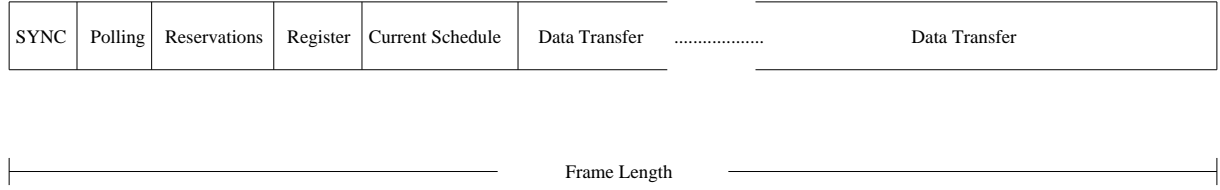


Figure 1: FDTMA Frame Structure

A frame is logically subdivided into phases (see figure). The available phases have the following function:

SYNC : In the base station a timer runs with the fixed frequency t_f . If the timer expires, a new frame starts. The base station transmits a short synchronization burst. This is a simple but effective method to achieve clock synchronization between all stations.

Polling : The base station maintains a list of all currently registered rt-clients (LARTC) in the picocell. During the polling phase the base station sends a short separate polling packet to every registered rt-client (all packets are sent in sequence). Every packet contains the station address of the rt-client, a checksum and two data fields: the first field gives the total number of stations to be polled, the second field gives the relative position of the current rt-client within the list of polled stations. With these two fields the rt-client can exactly determine the beginning of the next (reservation) phase and also the exact position within the reservation phase, during which the rt-client transmits its reservation packet (all reservation packets have the same fixed length). The last is determined by the relative position of the rt-client within the list of polled stations. If needed, the list of polling packets can be sent twice in order to increase the probability of proper reception.

Reservations : Every rt-client who receives a correct polling packet during the previous phase first determines at which time it may transmit its reservation packet, i.e. it first determines the start of the reservation phase and then determines its exact position within this phase. The reservation packet is of one of the following types: an alive-packet, a long-term-reservation- (LTR) packet, a short-term-reservation- (STR) packet, an unreserve-packet and an unregister-packet. The rt-client responds with an alive packet if it has no packets to transmit within this frame. If the client responds with an LTR packet, it wants to reserve a fixed number of k ($k \leq N$) slots in every i -th ($i \geq 1$) frame for exclusive use (dependent on the actually used scheduling scheme other parameters can be used). The

base station checks immediately whether the reservation can be accepted or not. The result is transmitted within the Current-Schedule-phase. The requested slots are assigned already within this frame (if the request is accepted). If the rt-client responds with an STR packet, it wants to reserve a number of slots within this frame. If it has already a LTR and it requests less than the reserved number of slots, the overlapping slots are distributed to other rt-clients (as long as there are unfulfilled requests) or marked as free (if all rt-clients are satisfied). If it wants more slots the base station assigns at least the reserved number of slots, but if other rt-clients give up slots it can get more. If the mobile responds with an unreserve-packet, it releases an existing LTR. If it responds with an unregister-packet the mobile is deleted from LARTC and will not be polled in the future.

Register : This is a single slot of fixed length, to which mobiles access with a contention based protocol for registering as a rt-client at the base station, i.e. to be included in the LARTC and to be polled in every frame.

Current Schedule : On the basis of the information gained in the polling and reservation phase of this frame and already existing reservations the base station computes a current schedule, i.e. an assignment of data slots to stations which can transmit in these slots for the actual frame. This schedule is broadcasted by the base station to all mobiles (if needed it is sent twice in order to increase the probability of proper reception at all stations). After reception any rt-client knows, which slot it can use exclusively for transmissions. Additionally, all stations (rt-clients and nonrt-clients) know which slots are free. If there are free slots the base station tries to spread these in an equidistant fashion and the first slot in the data phase will be a free slot. This should help to minimize access delay for best effort packets from nonrt-clients and / or rt-clients³.

Data Transfer : Every rt-client knows from the schedule in which slots it may transmit packets exclusively and which slots are free. A mobile may

³The tradeoff is the loss of exact jitter control. For a connection using slots in every frame the maximum jitter may be a little less than two frame lengths.

transmit its packets to all available stations including the base station if it serves as forwarder. If requested the receiver can send a short MAC layer acknowledgement. The free slots are accessed by all mobiles (rt-clients and nonrt-clients) using a contention based MAC protocol, e.g. slotted ALOHA together with a proper backoff algorithm. In principle rt-clients should make use of reservations, but if a packet arrives shortly after reservation phase and there are free slots it may be feasible to try to transmit the packet immediately.

The protocol framework described so far is flexible enough to implement a wide range of concrete protocols and packet scheduling policies for guaranteeing time and bandwidth requirements. Packet scheduling decisions need to be done in every rt-client and in the base station. The base station must also decide what happens if there are more short term reservations than slots. Different selection disciplines are possible. A very important future step will be the analytical and numerical evaluation of different selection disciplines and packet scheduling policies. I will give some short examples of how different packet scheduling policies can be implemented in the described protocol framework:

global EDF packet scheduling : This discipline can be implemented the following way: every rt-client transmits two numbers to the base station within the reservation phase: the first number gives the number of packets which will miss their deadlines, if they are not transmitted within this frame, the second number gives the total number of packets ready for transmission. The base station schedules the critical packets first. If there are more slots available, the remaining slots are used for noncritical packets. The assignment of free slots to the available stations can be done based on different policies. If there are more critical packets than slots available, the base station has to schedule a subset of the critical packets for transmission. The other packets must be discarded by the stations. It must be specified, how this subset is determined: one can assign all stations a slightly smaller number of slots than necessary or one can assign some stations no slots while allowing other stations that all their critical packets are transmitted. This last policy minimizes the number of stations that miss their deadlines.

local EDF packet scheduling : This discipline can be implemented easily locally. Every rt-client makes a LTR, applies the simple schedulability criterion given in [15] and accepts connections until the clients share of the total available bandwidth is completely used. Additionally the client makes use of rt-client.

strict TDMA : Every mobile makes a LTR and in every

frame he exactly requests the number of reserved slots with an STR packet.

Rate monotonic packet scheduling : As mentioned above, every rt-client sends two numbers i and k with its LTR packet. With these numbers the rt-client requests for k slots in every i -th frame. Based on this information and assuming that the deadline of each packet is equal to its period the simple schedulability test for the rate monotonic approach (described in [15]) can be applied.

Highest priority packet scheduling : This general discipline can be implemented as follows. Consider the number of available priority levels is p . Every station transmits a *priority allocation vector* consisting of binary numbers (one bit per priority level) during the reservation phase. In this vector the rt-client sets a bit for every priority level, for which he has a packet to transmit. The base station determines the transmission schedule based on the priority allocation vectors of all stations, always selecting high priority packets first. This approach is only efficient for a small number of priority levels due to the overhead needed for the priority allocation vectors.

The performance of an FTDMA based system, measured as throughput, schedulability, mean delay, and so on, is determined by a set of different parameters, which I will investigate in the future:

- the frame size t_f : this time defines the maximal resolution in time, i.e. in order to obtain a fine granularity in time t_f must be small. But with decreasing t_f the overhead increases.
- the number of slots per frame and the slot size.
- the maximum number of rt-clients within a cell.
- the scheduling discipline used in the mobiles and in the base station, which should take into consideration the media characteristics.
- the selection discipline for distributing unused slots to stations which want to have more slots than reserved.
- redundant transmission of control and signalling packets in order to increase the probability of proper reception. This is dependent from the error characteristics of the medium.
- system parameters like overhead, modem switching times,...

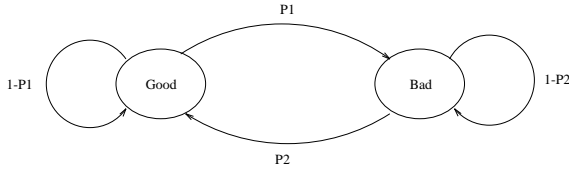
5 LINK-STATE-DEPENDENT SCHEDULING

In this section I introduce the basic idea of link state dependent scheduling and give a small example of how

this principle can be applied to existing scheduling policies (in this case EDF scheduling).

Consider a wireless picocell with a number of mobiles. A mobile A transmits packets to mobiles M_1, \dots, M_n . We assume that there is a separate logical channel between A and each mobile M_i . We assume that the error characteristics of all channels are statistically independent. As discussed in 2.1 there are a lot of error sources and the error characteristics are assumed to be nonstationary.

For modeling the error characteristics a simple model is widely used [22],[23]: consider a two state markov chain (general: n-state) with the states named *Good* and *Bad*.



Every state is assigned a specific bit error rate (BER), P_g in the good state, P_b in the bad state ($P_g \ll P_b$). All bit error rates are independent; they depend on the frequency and coding scheme used. The time is divided into slots of fixed length. A transition from one state to another occurs only at the beginning of a slot. If the channel is in good state, it changes to bad state at the beginning of a slot with probability P_1 and it stays in good state with probability $1 - P_1$. If the channel is in the bad state, it changes with probability P_2 to the good state, with probability $1 - P_2$ it stays in bad state.

From this description it can readily be determined that the time the channel stays in the bad state is a geometric random variable with mean $1/P_2$ (this is the mean length of an error burst) and that the mean BER is

$$\frac{P_2 P_g + P_1 P_b}{P_1 + P_2}$$

The mobile observes every channel and assigns one of the two states good and bad. This assignment can be based on information about signal strength or signal quality (e.g. some ARLAN radio modems have a received signal strength indicator), successful or missing acknowledgements, the occurrence of different error types (errors in the packet header, wrong checksum, loss of synchronisation) and so on. Good criteria need to be developed.

In [3] it was shown that a mobile can increase its throughput, if it postpones a packet transmission on a channel in bad state as long as possible and proceeds with transmission of packets on good channels, hoping that the channel quality on the bad channel changes to a good quality meanwhile. This approach is called *link state dependent scheduling*. The authors of [3] have shown experimentally that the efficiency of transmission increases significantly compared to round-robin packet scheduling. They identified head-of-line-blocking to be the main reason for performance degradation in case of round-robin

packet scheduling. Now this approach need to be applied to packet scheduling algorithms which are capable of guaranteeing service in a way, that no guarantees are violated.

It is determined by the underlying packet scheduling algorithm, whether and how link state dependent packet scheduling can be implemented. I will show as an example in the next section, how the EDF scheduling scheme can be adapted to link state dependent scheduling.

I expect the most gain in efficiency in environments with errors due to fast fading and noise, i.e. with error types which can change their characteristics very fast. In environments with errors mainly due to slow fading I assume that there is no significant gain in efficiency.

5.1 Link State Dependent EDF-Scheduling

We consider only a simple case. Let $\Gamma = \{\tau_1, \dots, \tau_n\}$ be a set of packet sources. Let the following assumptions hold (see [15]):

- time is slotted, the length of a time slot is L
- all packet sources generate packets in a cyclic fashion with period T_i , expressed in multiples of L
- For every packet source all generated packets have the same length C_i , expressed in multiples of L .
- A deadline D_i is associated with every packet, which is equal to the period T_i .
- At the beginning of a slot a packet can be preempted by another packet.
- All packet sources have the same starting time 0.
- there are no precedence relations between packets

Packets are scheduled by the EDF scheduling rule, i.e. always the packet with the shortest deadline is chosen. New arriving packets can preempt transmission of currently active packets. If all the above conditions are true, then the following holds [15]:

1. If any scheduling algorithm can find a feasible schedule, i.e. a schedule where no packet miss its deadline, then the EDF algorithm can.
2. A necessary and sufficient condition for schedulability of a set of packet sources Γ is:

$$\sum_{i=1}^n U_i := \sum_{i=1}^n \frac{C_i}{T_i} \leq 1$$

where U_i is the resource utilization generated by packet source i . If we define $T = \text{LCM}\{T_1, \dots, T_n\}$ (Least Common Multiple) and $I_i = T/T_i$ (Number of packets generated by

source i within time $[0, T)$) this can be rewritten as

$$\sum_{i=1}^n I_i C_i \leq T$$

Consider a set of packet sources Γ which do not fully utilize the shared resource (i.e. $\sum_{i=1}^n U_i =: U < 1$). The basic idea is to assign each packet source a new “virtual” packet length $C'_i \geq C_i$. The difference between C'_i and C_i can be used to postpone packet transmission if the link is in a bad state. Let

$$C := \sum_{i=1}^n I_i C_i$$

and $R := T - C$ the number of remaining slots. It is then possible to increase uniformly the packet length for every packet of every packet source within $[0, T)$ by the amount of

$$k := \left\lfloor \frac{R}{T} \right\rfloor$$

(therefore $C'_i = C_i + k$ for all i) without missing a deadline⁴. One problem with this scheme is, that the utilization increase for packet sources with low utilization is very large as is the “postponing gain”. On the other hand packet sources with a high utilization receive a relatively small postponing gain.

A second strategy for assignment of remaining slots allocates additional slots in order of decreasing utilizations (utilization monotonic approach): Let the packet sources be numbered in order of decreasing utilizations, i.e. the relation

$$U_1 \geq U_2 \geq \dots \geq U_n$$

holds. Then let

$$\begin{aligned} R_0 &:= R \\ k_i &:= \left\lceil R_{i-1} \frac{U_i}{U} \right\rceil, \quad (1 \leq i \leq n) \\ R_i &:= R_{i-1} - k_i, \quad (1 \leq i \leq n) \end{aligned}$$

and finally let $C'_i := C_i + k_i$. It is clear, that always $R \geq R_i \geq 0$ and $\sum_{i=1}^n k_i = R - R_n$ holds, so all deadlines are met. However, this scheme is not fair to packet sources with the same utilization: consider packet sources τ_1, τ_2, τ_3 , each with $C_i = 1$ and $T_i = 8$. Then we have $R = 5, k_1 = 2, k_2 = 1, k_3 = 1$ and $k_1 + k_2 + k_3 = 4$. So the remaining slots are not distributed equally and the resource is not fully utilized. In this case a second run of the algorithm, but now in order of increasing utilizations, will give τ_3 a second slot.

A third strategy uses a constant s with $0 < s < 1$ and assigns a new packet length by

$$C'_i := \left\lceil \frac{1}{1-s} C_i \right\rceil$$

In this case also all deadlines are met if the relationship $\sum_{i=1}^n U_i \leq 1 - s$ holds.

One can think of several other strategies for allocating remaining slots in order to postpone packets. One of my near future goals will be the numerical and analytical evaluation of such strategies.

From an implementation point of view the modified EDF algorithm can be implemented using a priority queue ordered by deadlines. The scheduler always transmits the first packet in the queue. If the link needed for packet transmission is in a bad state and there is no remaining time for postponing the packet, the scheduler starts transmission of the first packet. If there is remaining time for postponing the packet, the scheduler traverses the priority queue until it finds a packet which uses a good link or it has all packets inspected. In the latter case (all needed links are in a bad state) the scheduler waits the remaining time for postponing the first packet before he starts transmission. So the scheduler becomes non work-conserving. If the scheduler finds a good link, it starts transmission of the corresponding packet until the remaining time for postponing the first packet has expired. Then the packet for the good link is preempted and transmission of the first packet begins.

As already mentioned, the efficiency of this scheme depends on how exactly the channel state is determined and on the error characteristics of the real wireless channel. In the near future I want to investigate these questions on the basis of simulations.

In this section the link state dependent scheduling approach is only applied to a very simple case, in which all packet sources are cyclic. A very important future step will be to extend this approach to sporadic messages and to further apply this approach to other packet scheduling algorithms (e.g. rate monotonic packet scheduling).

6 CONCLUSIONS

I have presented some ideas regarding different elements of an architecture for implementation of wireless industrial LANs, namely an overall system architecture, a description of data link layer services, which reflects the users needs in defining timing and bandwidth requirements, a proposal for a MAC protocol capable of implementing a wide range of different concrete protocols and packet scheduling policies and an approach for modifying packet scheduling algorithms in order to accommodate to the nonstationary error characteristics of wireless media.

Clearly these are only some elements of a wireless industrial LAN. The remaining design areas include: finding a proper mapping between abstract data link layer

⁴ $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x , $\lceil x \rceil$ denotes the smallest integer greater than or equal to x .

services and the FTDMA scheme, finding a mapping between link layer services and ATM, mobility management and handover problems, schemes for (adaptive) redundant coding for error detection and recovery in the physical layer, link layer protocols, application layer issues and so on.

My research in the next future will be concerned with the refinement and analytical/numerical evaluation of the presented approaches, namely the FTDMA approach and the link state dependent scheduling approach separately and closely integrated., and also in finding a scheme for efficient and smooth integration of the presented ideas.

References

- [1] K. Arvind, Krithi Ramamritham, and John A. Stankovic. A local area network architecture for communication in distributed real-time systems. COINS Technical Report 91-04, Dept. of Computer and Information Science, University of Massachusetts at Amherst, January 1991.
- [2] David F. Bantz and Frederic J. Bauchot. Wireless lan design alternatives. *IEEE Network*, 8(3):43ff, 1994.
- [3] P. Bhagwat, P. Bhattacharya, A. Krishna, and S. Tripathi. Enhancing throughput over wireless lans using channel state dependent packet scheduling. In *Proc. IEEE INFOCOM 96*. IEEE, 1996.
- [4] Kenneth L. Blackard, Theodore S. Rappaport, and Charles W. Bostian. Measurements and models of radio frequency impulsive noise for indoor wireless communications. *IEEE Journal on Selected Areas in Communications*, 11(7):991–1001, September 1993.
- [5] Klaus David and Thorsten Benkner. *Digitale Mobilfunksysteme*. Informationstechnik. B.G. Teubner, Stuttgart, 1996.
- [6] Jean-Dominique Decotignie and Patrick Pleineveaux. A survey on industrial communication networks. *Ann. Telecomm.*, 48(9):435ff, 1993.
- [7] Martin DePrycker. *Asynchronous Transfer Mode - The solution for Broadband-ISDN*. Prentice Hall, New York , London, second edition, 1994.
- [8] Antonio DeSimone and Sanjiv Nanda. Wireless data: Systems, standards, services. *Wireless Networks*, 1(3):241ff, 1995.
- [9] DIN - Deutsches Institut f'ur Normung, Beuth Verlag Berlin. *DIN 19245 Teil 1 - PROFIBUS: "Ubertragungstechnik, Buszugriffs- und "Ubertragungsprotokoll, Dienstschnittstelle zur Anwendungsschicht, Management*, April 1991.
- [10] D. Duchamp and N.F.Reynolds. Measured performance of wireless lan. In *Proc. of 17th Conf. on Local Computer Networks*, Minneapolis, 1992.
- [11] David Eckhard and Peter Steenkiste. Measurement and analysis of the error characteristics of an in-building wireless network. In *Proc. of ACM SIGCOMM'96 Conference*,, pages 243–254, Stanford University, California, August 1996.
- [12] Ralf Holtkamp, Jean-Pierre Ebert, Adam Wolisz, and Louis Ramel. A distributed media access control (dmac) for wireless atm networks. In *Proc. 5th Telecommunication System Conference*, Nashville, March 1997.
- [13] Ivan Izikowitz and Michael Solvic. Industrial needs for time-critical wireless communication & wireless data transmission and application layer support for time critical communication. http://www.faps.uni-erlangen.de:1200/persons/stoeckel/olchfa/euro_arc.ps.gz, 1994.
- [14] Mark J. Karol, Z. Liu, and K.Y. Eng. An efficient demand-assignment multiple access protocol for wireless (atm) networks. *Wireless Networks*, 1(3), 1995.
- [15] C.L. Liu and J. Layland. Scheduling algorithms for multiprogramming in a hard real-time environment. *Journal of the ACM*, 20(1):46–61, 1973.
- [16] Philip Morel and Alain Croisier. A wireless gateway for fieldbus. In *Proc. Sixth International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 95)*, 1995.
- [17] Philip Morel and Jean-Dominique Decotignie. Integration of wireless mobile nodes in map/mms. In *Proc. 13th IFAC Workshop on Distributed Computer Control Systems DCCS95*, 1995.
- [18] Raif O. Onvural. *Asynchronous Transfer Mode Networks - Performance Issues*. Artech House, Boston, London, second edition, 1995.
- [19] K. Pahlavan and A.H. Levesque. *Wireless Information Networks*. J. Wiley and Sons, 1995.
- [20] Dietmar Petras. Medium access control protocol for wireless, transparent atm access. In *IEEE Wireless Communication Systems Symposium*, Long Island, NY, November 1995. IEEE.
- [21] John Porter and Andy Hopper. An atm based protocol for wireless lans. ORL Technical Report 94.2, Olivetti Research Laboratories, 1994.
- [22] F. Swarts and H.C. Ferreira. Markov characterization of digital fading mobile vhf channels. *IEEE Transactions on Vehicular Technology*, 43(4):977–985, November 1994.
- [23] H.S. Wang and N. Moayeri. Finite state markov channel - a useful model for radio communication channels. *IEEE Transactions on Vehicular Technology*, 44(1):163–171, February 1995.